Data Product Manager Nanodegree

Applying Data Science to Product Management
Final Project: Developing an MVP Launch Strategy for a Flying Taxi Service

TABLEAU PUBLIC LINK

https://public.tableau.com/app/profile/xabier5536/viz/taxi_rides1/dist_dur_ratio_drop_off?publish=yes

Welcome to your first week at Flyber


Aybar

In this project, you will apply the skills acquired in this course to create the MVP launch strategy for the first flying car taxi service, Flyber, in one of the most congested cities in America -- New York City.

You are responsible for bringing the first flying car taxi service to market by analyzing data and building a product proposal.

You will need to use the SQL workspace provided in the Classroom, and Tableau Public, in order to successfully complete the project.

You'll present your answers, findings, and insights in the Answer Slides found in this deck. Feel free to include any additional slides, if needed.

Section 1: Data Exploration

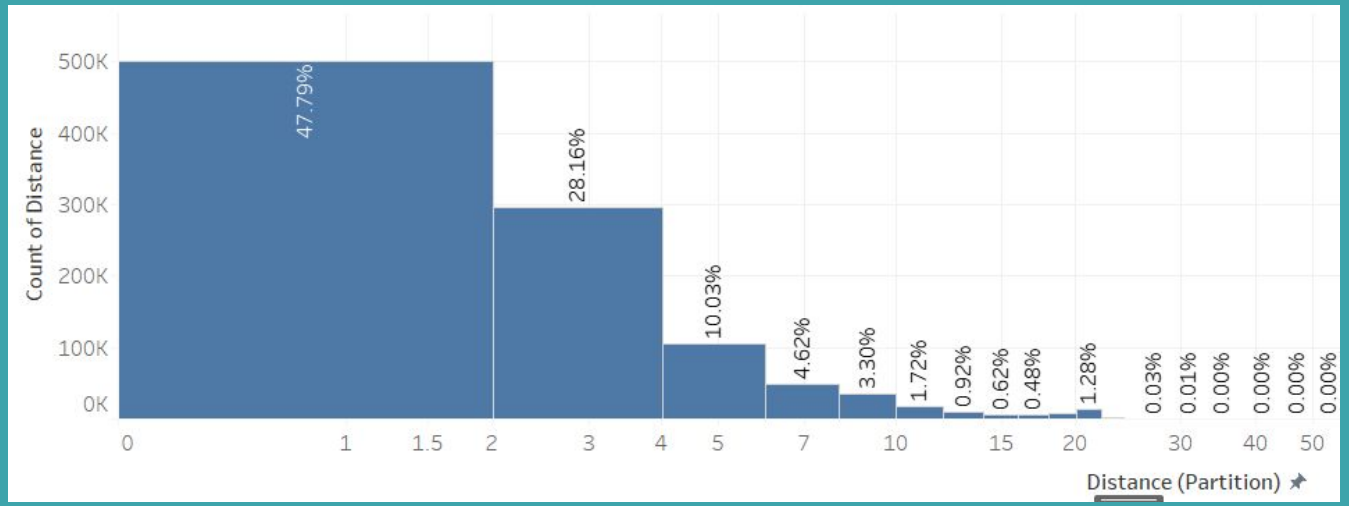
Back to the basics of product management, identify your customer and their pain points:

- What are taxis used for?
- What are the characteristics of the users that leverage them?
- What are existing pain points with taxis?
- What are the existing pain points with digital ride-sharing services?

Answer Slide

- **What are taxis used for?**

→ In general the taxis are used for short rides. More than 75% of the rides are shorter than 5 miles



Answer Slide

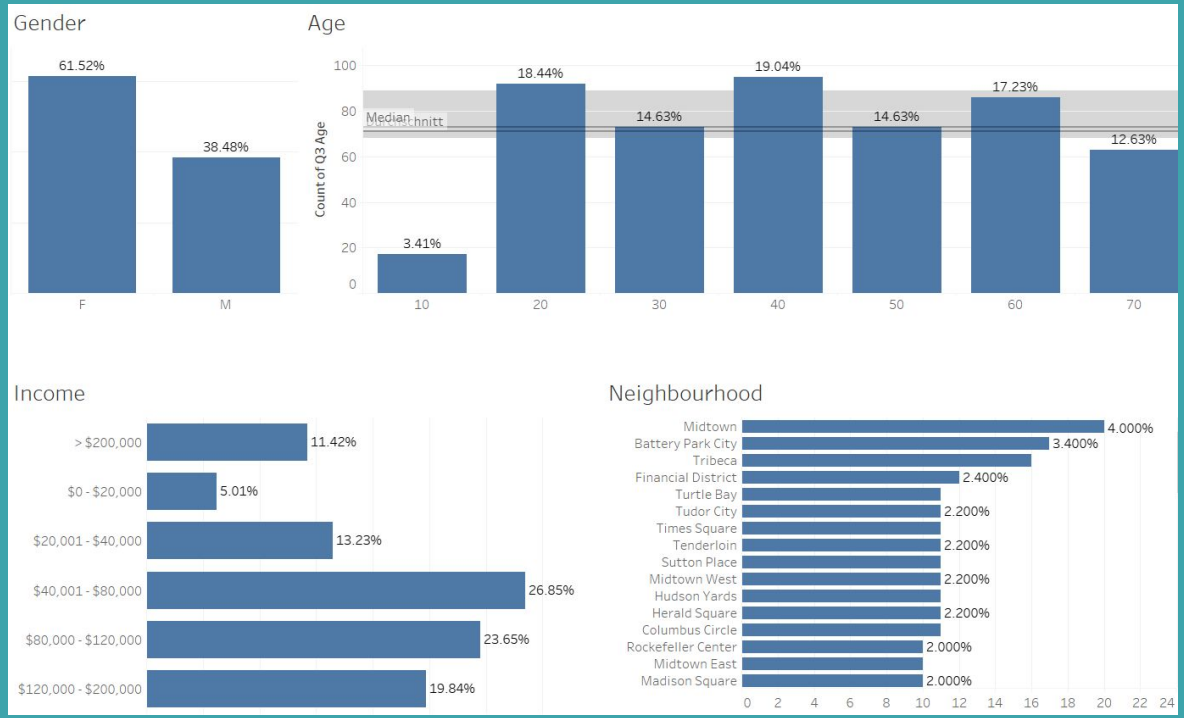
- **What are the characteristics of the users that leverage them?**

  ⟶ In general terms, the characteristics of the users are the following:

  - Female
  - 20-30 or 40-50 years old
  - Income range: 40k-80k$
  - Lives in Midtown neighbourhood

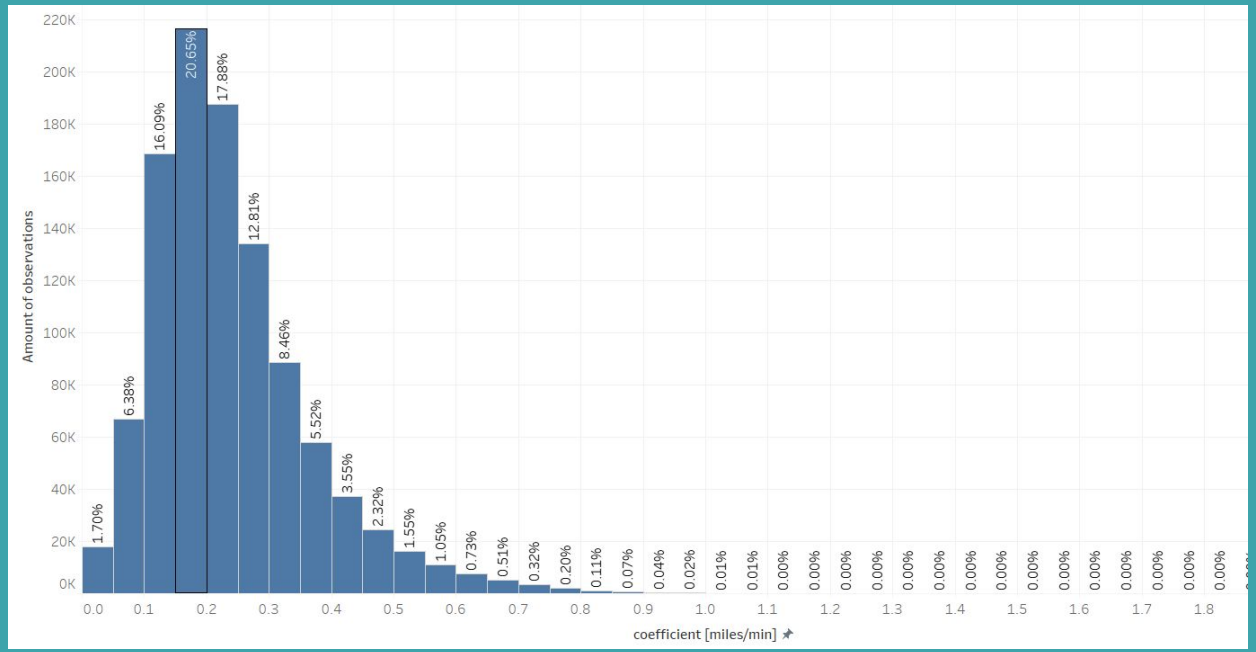
Answer Slide

- **What are the characteristics of the users that leverage them?**



Answer Slide

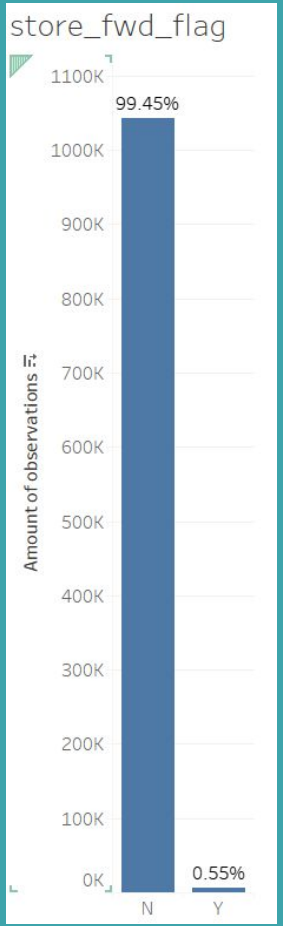
- **What are existing pain points with taxis?**
  → Most of the rides need too long to cover a mile. Almos 45% of the rides require 5 minutes or more to cover 1 single mile.



Answer Slide

- **What are the existing pain points with digital ride-sharing services?**

  → Most of the data is sent without storing it. This is a pain point in the data collection process.



What user improvements do you hypothesize a flying taxi service would have over the existing state of taxis today?

What market improvements do you hypothesize a flying taxi service would have the existing taxi service industry & physical road infrastructure today?

Answer Slide

- Regarding the user improvement, I hypothesize that a flying taxi will be faster, specially within the city. Additionally, it will make a big difference for the short rides, which happen to take to long very often.

- The biggest impact would be the reduction of traffic in the regular taxi services, due to the absorption of rides by the flying taxis. On the other hand, a new business like the flying taxis could generate new related businesses that do not currently exist.

Upload this dataset into Tableau Online.

Ensure the fields are parsed correctly; field headers are included in the first row of the CSV.

Let’s begin exploration!

Acquire a high-level understanding of the granularity and scope of the dataset, to inform the basis for your analyses:

- How many records are in the dataset
- What does each record represent?
- What is the primary key?
- What date range is your dataset bound to?
- What are the geographical bounds of this dataset? Is it limited to Manhattan, or is Brooklyn, Queens, Staten Island, the Bronx, and New Jersey included? Where are most of the data points centralized at? Are there outliers?

Answer Slide

- In the *taxi rides* file there are 1,048,468 records.
- Each record represents a ride.
- The primary key is the “Id” column, which is the identification of the ride.
- The date are within *01/01/2016 00:00:00* to *30/06/2016 23:59:00*.
- The geographical bounds are not limited just to Manhattan, Brooklyn, Queens, Staten Island and the Bronx are also included. The highest density of dropoffs takes place in Manhattan and Brooklyn.
- Yes, there are outliers, since there is a dropoff in Toronto and some more in the sea.

You notice that the dataset does not contain explicit data points out-of the-box, we'll need to enrich the dataset with relevant fields:

- You notice that ride price is not included, but figure it could be derived. Based on information about New York taxi prices gleaned from the internet, create a calculated field called `price` using the `duration`, `distance`, and `passenger count` fields.
- You hypothesize your target users will be those who take a relatively longer time getting to a destination that is relatively close, due to heavy traffic conditions and/or limitations to physical road infrastructure. To be able to analyze where this is happening, you will need to create a calculated field called `distance-to-duration ratio`.

Let’s understand the scope and distribution various dimensions within the dataset. Calculate the **average**, **median**, and the **first & second standard deviation of the mean** for the following measures:

- duration
- distance
- passenger counts
- distance-to-duration ratio
- price

Answer Slide

Duration

- Average = 962
- Median = 662
- Std1 = 5,853
- Std2 = 11,707

Distance

- Average = 3.422
- Median = 2.095
- Std1 = 4.382
- Std2 = 8:764

Passenger counts

- Average = 1.664
- Median = 1
- Std1 = 1.314
- Std2 = 2.628

Distance-to-duration ratio

- Average = 0.2405
- Median = 0.2132
- Std1 = 0.2374
- Std2 = 0.4748

Price

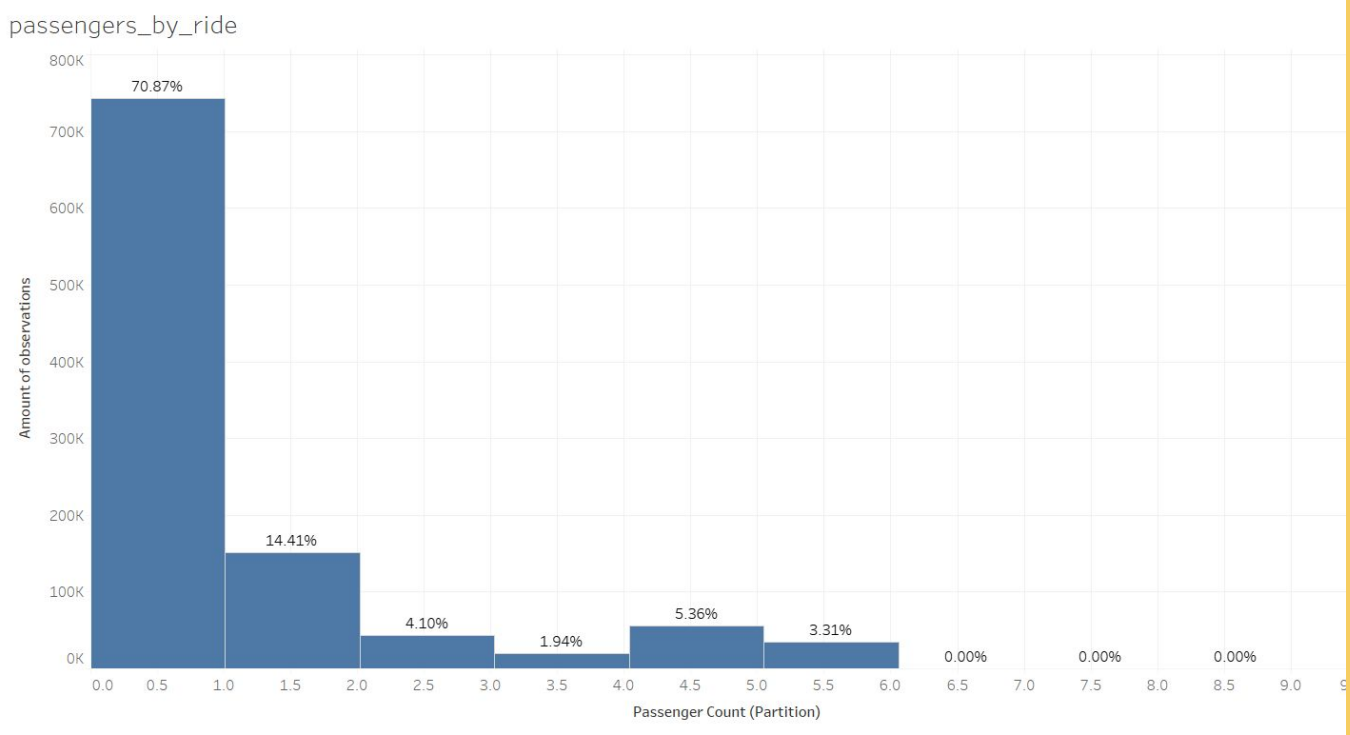
- Average = 51
- Median = 8
- Std1 = 577
- Std2 = 1,153

Flying cars may have to have to be a lower weight for efficiency & take-off. Or you may just decide to leverage mini-copters for your initial MVP.

Create a histogram that visualizes the number of total rides grouped by passenger counts to analyze the potential market volume of low passenger pickups (1-2 passengers).

Answer Slide

As we can see below 85.28% of the rides contain just 1 or 2 persons, so Flyber should focus on these kind of rides and go for the mini-copters, for efficiency reasons.



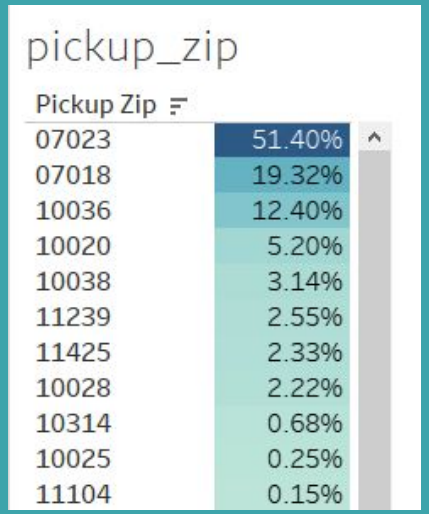
For the initial MVP launch (& most likely GA), we have a finite amount of monetary resources to build Flyber pick-up / drop-off nodes. We'll need to be strategic on where we'll place them:

- Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?
- Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?
- Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off?
- For any of the neighborhoods identified, are there any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off? What makes them suitable?

Answer Slide

- **Which neighborhoods/zip codes tends to experience a relatively higher density of pick-ups?**

  → The highest pick-up density is experienced in the 7023 and 7018 zip codes.



pickup_zip

Pickup Zip	
07023	51.40%
07018	19.32%
10036	12.40%
10020	5.20%
10038	3.14%
11239	2.55%
11425	2.33%
10028	2.22%
10314	0.68%
10025	0.25%
11104	0.15%

Answer Slide

- **Which neighborhoods/zip codes tends to experience a relatively higher density of drop-offs?**

  → The highest drop-off density is experienced in the 7023 and 7018 zip codes.

dropoff_zip

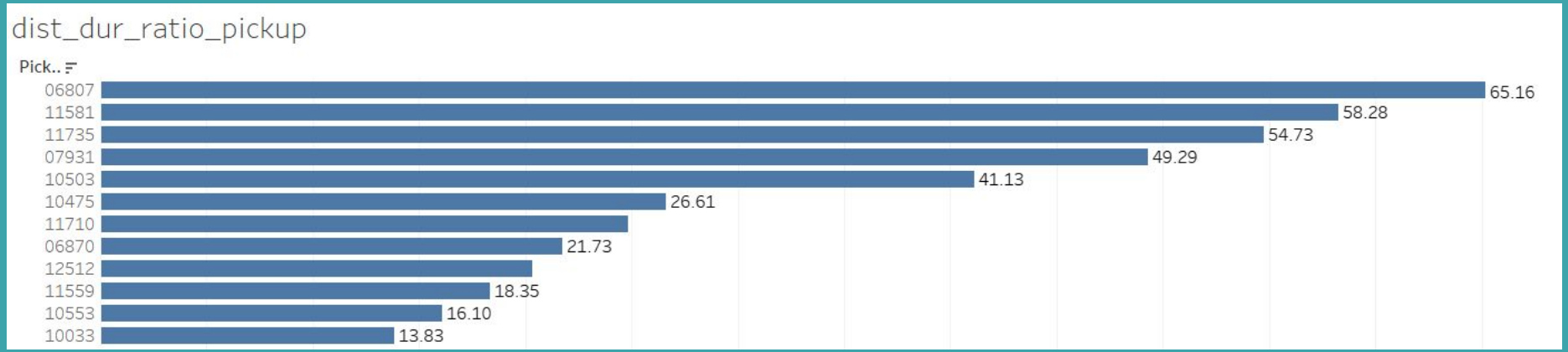
Dropoff Zip	
07023	48.81%
07018	18.48%
10036	13.45%
10020	4.51%
10038	4.31%
10028	2.54%
10314	2.32%
11239	1.74%
11425	1.59%
11104	0.48%

Answer Slide

- **Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on pick-up?**

→ For rides that have a duration-to-distance ratio lower than 150 minutes/mile the hughes values happen in the following zip code:

- **06807 → 65.16 minutes/mile on average**

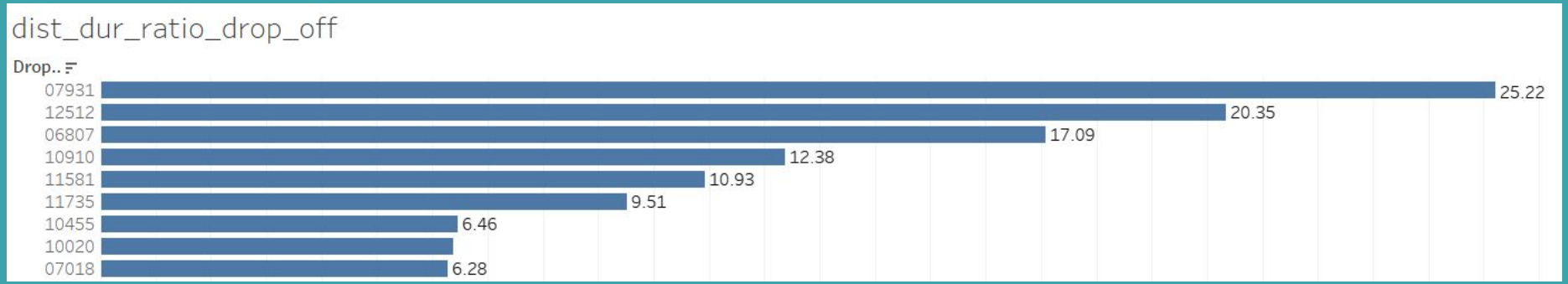


Answer Slide

- **Which neighborhoods/zip codes tends to have the highest duration-to-distance ratios, based on drop-off?**

→ For rides that have a duration-to-distance ratio lower than 150 minutes/mile the hughes values happen in the following zip code:

- **07931 → 25.22 minutes/mile on average**



Answer Slide

- **For any of the neighborhoods identified, are there any potential areas within the neighborhood that are optimal for flying taxi pick-up / drop-off? What makes them suitable?**

→ Yes, Midtown because the 75% of the users in this neighborhood are will to share the taxi, so the price of the ride will be higher for the same distance/duration.

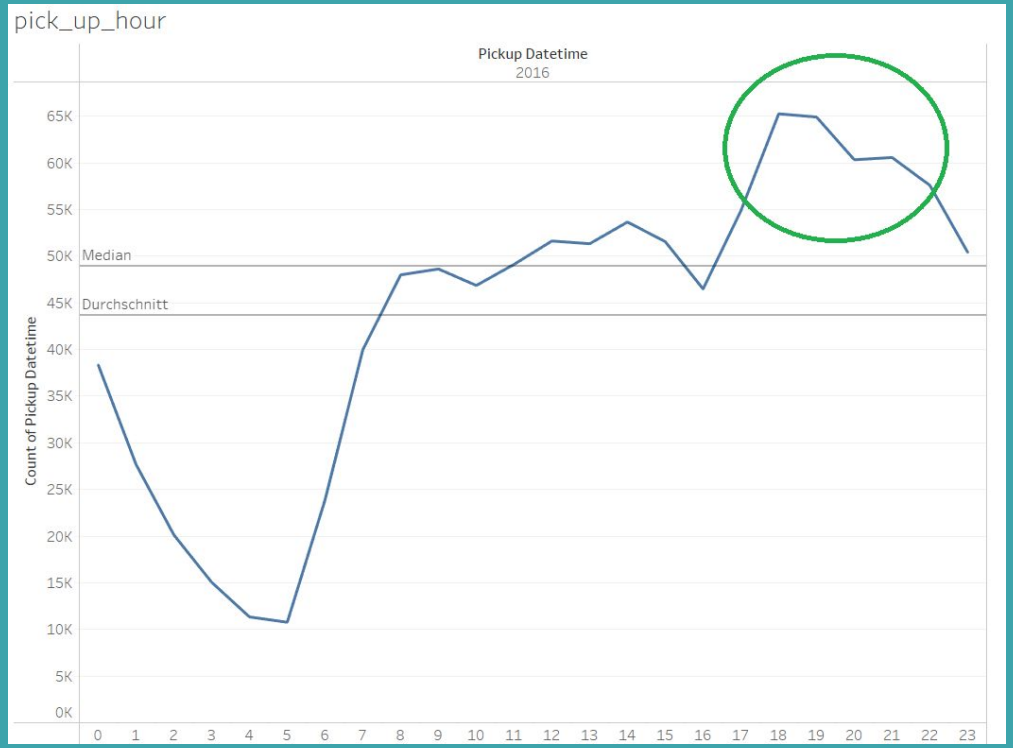
It may not make operational sense to have the service running 24/7, for now.

- What times throughout the day experience relatively higher volumes of ride pick-ups?
- What days throughout the week experience relatively higher volumes of ride pick-ups?
- Pinpoint any periods throughout the year that experience trend fluctuation or seasonality around ride pick-up volumes. This will help us in our post-launch analyses to determine if any spikes or dips were influenced by seasonality or through actual feature adoption/regression.

Answer Slide

- **What times throughout the day experience relatively higher volumes of ride pick-ups?**

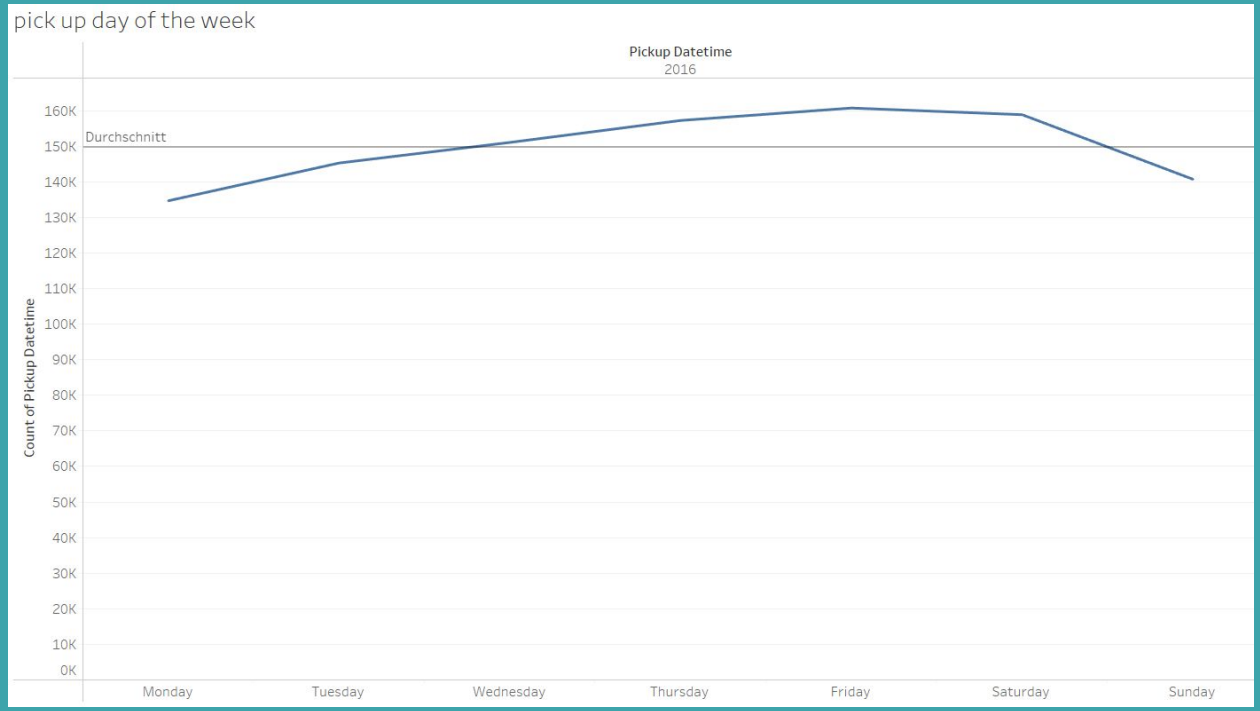
From 18h to 22h there are higher volumes of ride pick-ups.



Answer Slide

- **What days throughout the week experience relatively higher volumes of ride pick-ups?**

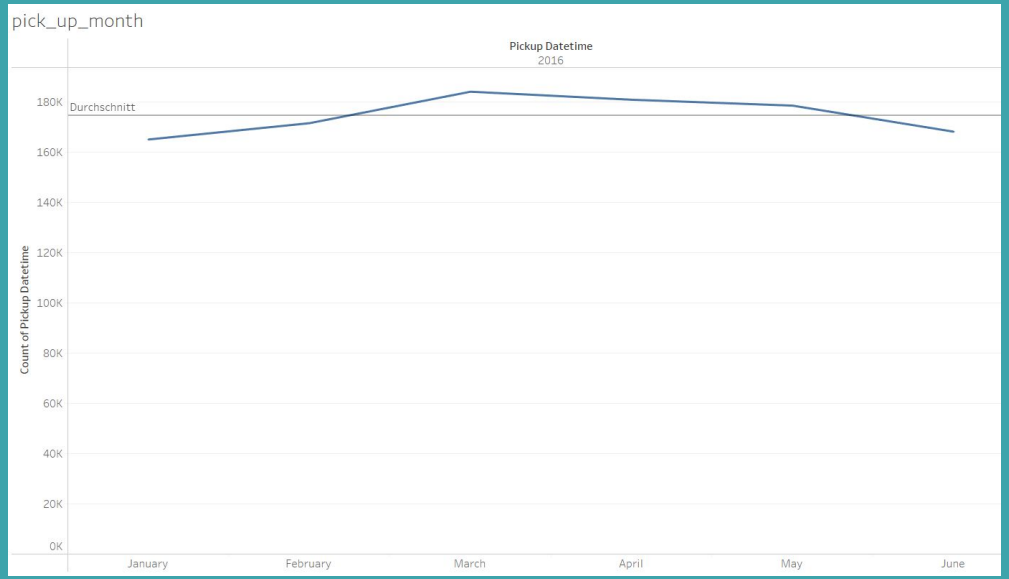
The days of the week with higher volumes of ride pick ups are Friday and Saturday.



Answer Slide

- **Pinpoint any periods throughout the year that experience trend fluctuation or seasonality around ride pick-up volumes. This will help us in our post-launch analyses to determine if any spikes or dips were influenced by seasonality or through actual feature adoption/regression.**

As we can see in the picture on the right, in March and April there are slightly more rides than in the other months.



You and the user research team ran a quantitative survey on existing taxi and/or rideshare users in New York City to determine sentiment around potentially using a flying taxi service.

Dive into the survey results dataset in order to extract insights from explicit feedback.

Upload this dataset into Tableau Online or a SQL database (the classroom contains a workspace with the data for you as well).

Ensure the fields are parsed correctly, field headers are included in the first row of the CSV.

Question schema:

Q1 - What is your email?
Q2 - What gender do you identify as?
Q3 - What is your age?
Q4 - What is your annual income? (income bands)
Q5 - What neighborhood do you reside in?
Q6 - Do you currently use taxis? (Y/N)
Q7 - Do you currently use ridesharing services? (Y/N)
Q8 - Would you use a flying taxi service, if such a concept existed? (Y/N)
Q9 - If yes to Q8, how much would you be willing to pay per mile for such a service? (USD)
Q10 - If no to Q8, what is the reason?

To inform our future product marketing efforts, we'll want to extract the following:

- Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?
- What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?
- What is the different personas/segments of negative sentiment towards not using a flying taxi car service?

Answer Slide

- **Is there an inclination of better Flyber adoption based on gender, age, income level, or neighborhood of residence?**

→ Yes, 64% of the users are female.

→ Regarding the age, most of the users are younger than 20 (21.85%) or between 30-40 years old (19.04%).

→ The neighbourhoods with more users are Midtown (4% of the users), Battery Park City (3.4%) and Tribeca (3.2%).

→ Attending to the location, there are some neighbourhoods, that even do not use this service (within men) like West Harlem and other with very low rate, like East Village among others (0.521%).

Answer Slide

- **What is the distribution of potential price per mile based on gender, age, income level, and neighborhood of residence?**

→ The age, gender, income level and neighborhood of residence should not affect the price per mile. The only impact in the price will done by the duration of the ride and the people who share the taxi.

Answer Slide

- **What is the different personas/segments of negative sentiment towards not using a flying taxi car service?**

→ Men are not that prone to use flying taxi car services.

→ Within males, the age group uses taxi services the least is the group of people who are between 60-70 years old with a 9.38% of the users.

Hooray! End of Section 1.

You will complete Section 2 at the end of this course.

Please submit this file for review for Section 1.

Section 2: Proposal Synthesis

Identify a product objective for Flyber's launch. Your product objective will guide your KPIs, so identify what Flyber should optimize for. Your objective should be centered around one the following focus areas:

- User Acquisition
- User Engagement
- User Retention
- Profitability

Explain your reasoning. Include both why you feel your focus area is more relevant than the others for Flyber at this time of the product development cycle.

Product objective for Flyber launch

As this is a new service (it does not exist any other company in the market, that offers flying rides), we will first focus on acquiring users. After achieving this goal, we will need to further focus on profitability and user retention, and then, we will be able to focus on user engagement, so we can build a long term *Flyber-user* relationship.

Formulate 3-5 Key Performance Indicators (KPIs), to measure if the product is heading towards the right direction based on your objective

KPIs

1. Price by ride
2. Amount of rides per day
3. Downtime
4. Weekly customer acquisition increase
5. Customer satisfaction (experience satisfaction & recommendation likelihood

Create hypotheses around what thresholds your KPIs would need to hit in order to determine success

KPI thresholds

1. Price by ride → 80% of the rides more expensive than 10$
   Price>10$ → 0,8*Total_amount_rides

2. Downtime → lower than a 65% of the working hours
   Downtime<0,65*working_hours

3. Weekly customer acquisition increase → 10% increase the first month and 7% increase the second month and %5 increase the third and fourth months.
   1st week → current_customers*1,1< customers1
   2nd week → customers1*1,07< customers2
   3rd week → customers2*1,05< customers3

4. Customer satisfaction average → higher than 4 out of 5 (on average)
   AVG(Customer_rating)>4

As the product manager, you make decisions based on the insights you extract, we'll need to know the feature set we'll include in the MVP to measure viability, while keeping operational expenditure under control:

- What times/days of operation should the service run for?
- How many pick-up / drop-off nodes should we have?
- Where should the nodes be located?
- Should we initially use copters or homegrown hardware?
- Should the pricing be fixed or dynamic? At what rates?

Answer Slide

- **What times/days of operation should the service run for?**

→ Based on the hourly use of taxis, the Flyber service will work from 8h to 24h, two 8 hour shifts.

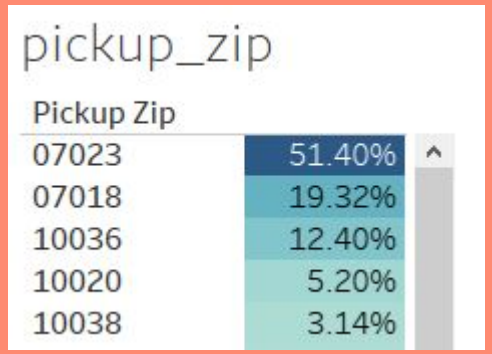
→ Even though the peak days are Thursday, Friday and Saturday, there will be Flyber service all the weekdays, as acquiring new customers is intended.

Answer Slide

- **How many pick-up / drop-off nodes should we have?**
- **Where should the nodes be located?**

→ It would be interesting to fix the pick-up nodes and to leave the drop-off points open to the customer needs (with a limit of 50 miles from the pick-up point).

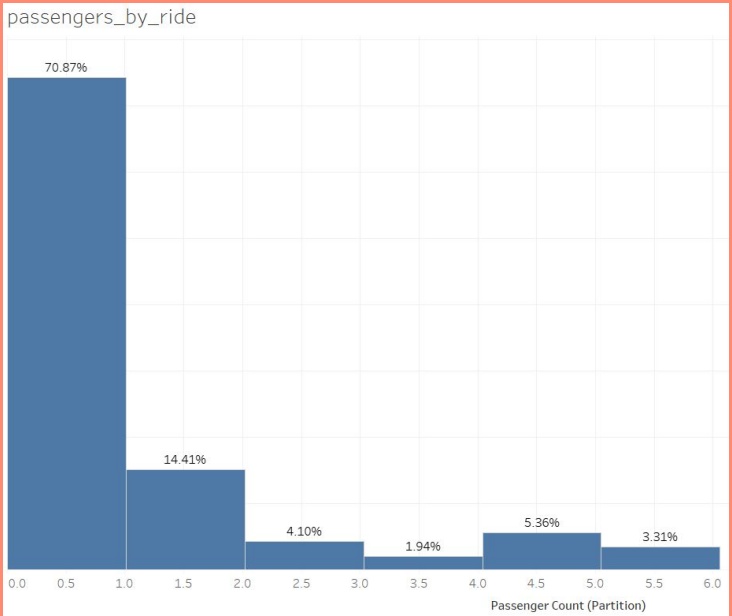
→ We will choose 5 pick-up points, the ones that represent about the 90% of the pick-ups for taxi rides:



Answer Slide

- **Should we initially use copters or homegrown hardware?**

→ As the taxi rides with 2 or less passengers represent the 85% of the rides, we will initially just use copters.



Answer Slide

- **Should the pricing be fixed or dynamic? At what rates?**

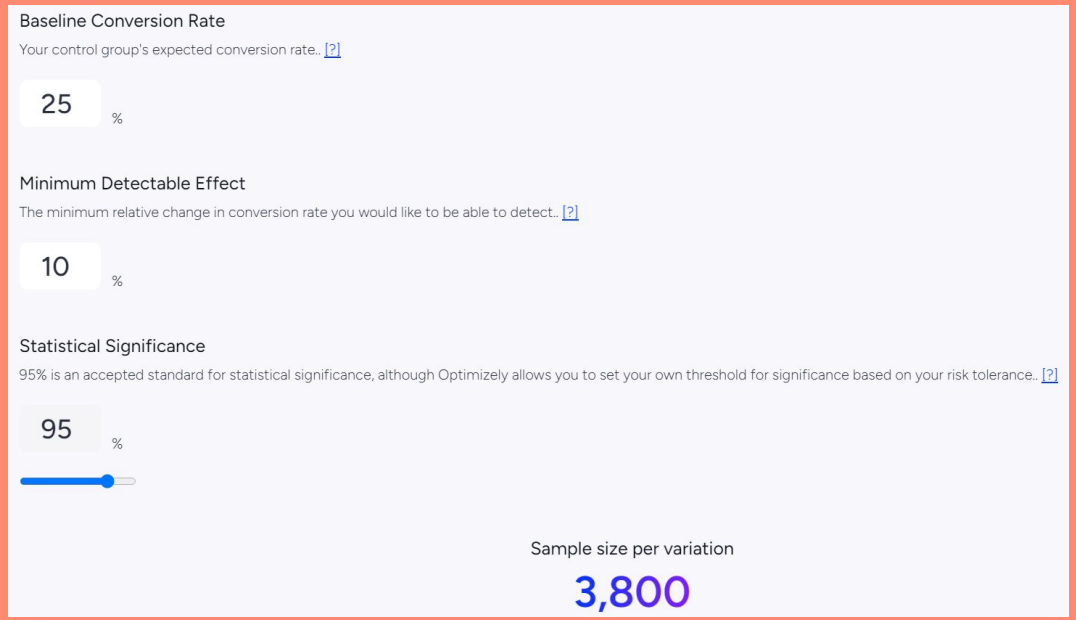
→ The price should be fixed for a combination of distance, duration and amount of passengers.

→ Price should not depend on time of the day, day of the week or copter availability. The customer should feel respected and that there is no speculation with the market situation (created by rush hours, bank holidays or other factors) that could generate a price rise to take advantage of the situation.

Determine the MVP sample size & time period allotted estimated to come to a conclusion on your hypotheses.

MVP sample size & hypothesis period

- Baselina conversion rate: 25%, confidence level: 95%, MDE=105.
- Population size: 3,800.
- Hypotheis period: 12 working days, including weekends (16 working hours/day, 5 rides/hour, 4 copters).



The conversion rate will be determined by the amount of people that uses the service among the ones that have a user account.

Create an instrumentation plan for the events you need collected and logged, in order to be able to physically measure your KPIs.

Instrumentation plan

These will be the events and properties belonging to the instrumentation plan:

- Event: rideIdentifier
  Definition: an id of the ride will be generated when the ride is booked
  Properties: ride_id, payment_time, price, payment_time, pickup_zipcode, dropoff_zipcode.
- Event: driverIdentifier
  Definition: triggers every time a ride is booked.
  Properties: driver_id, ride_id, copter_id.
- Event: copterIdentifier
  Definition: triggers every time a ride is booked
  Properties: copter_id, ride_id, copter_model
- Event: rideDuration
  Definition: triggers everytime a payment takes place.
  Properties: pickup_datetime, dropoff_datetime, ride_duration, ride_id.

Instrumentation plan

These will be the events and properties belonging to the instrumentation plan:

- Event: paymentInformation.

  Definition: triggers whenever a payment takes place.

  Properties: payment_id, price, payment_method, ride_id.

- Event: passengerInformation.

  Definition: triggers whenever a booking is booked in through the app.

  Properties: passenger_id, ride_id, people_amount, gender.

- Event: customerRating

  Definition: triggers every time after a ride finished/was paid.

  Properties: rating_id, customer_id, ride_id, driver_id, copter_id, timestamp.

Create a qualitative feedback survey questions for users after their ride, to further understand and optimize the product for future iterations.

Qualitative feedback survey

- All the survey questions can have a value from 1 to 5.

1. How would you rate the experience?
2. Would you recommend it to a friend?
3. Do you think the copter is better than a regular taxi?
4. How would you rate the distance-duration rate?

Summarize everything you have learned into your final proposal

- Identify the target population. Why did you select that target population? What are their pain points?
- Create a product proposal containing claim, evidence, estimated impact, and risks
- Claims should be backed by quantitative evidence, impact should assess market needs/benefits
- Risks involve any known unknowns that we'll still need to monitor post-launch
- State cross-functional stakeholder teams that will need to be involved

Answer Slide

- **Identify the target population. Why did you select that target population? What are their pain points?**

→ The target population will be people of any gender, slightly with more consideration to female of any age located in the 07023, 07018, 10036, 10020 and 10038, which represents about a 90% of the total pick-up in the taxi rides database.

→ The main pain point is the duration-to-distance rate which also impacts the price. Reducing the duration will be one of the main goals. By reducing the duracion at the same price, Flyber is generating a competitive advantage and at the same time more copter time availability, which can lead to a higher ride per hour rate.

Answer Slide

The Flyber will fulfill the need of the customer of faster rides to avoid traffic jam and a better user experience due to the absence of traffic chaos. Providing a “flying taxi” time will be saved and the same time, the cost related to the ride duration.

This product will have an impact in the market, changing the way of the customer of understanding the transport. As a consequence, the regular taxi business will lose clients, that will use the copter transportation.

As an estimation, 30% of the market share, that currently uses regular taxis will use copters. Along the implementation process it will be analized at what pace more copters should be put in service and if copters for more than two passenger are required and if they would be meaningful in terms of profit and customer acquisition.

Answer Slide

The main risks of the implementation will be related to infrastructures and the legal requirements/impediments to grow and escalate the business. Additionally, the customer mentality will play an important role, since maybe people do not trust on flying vehicle or maybe simply they are afraid of flying.
Another risk would be that any potential infrastructure barrier could impact the time saving in terms of ride duration reduction.

Answer Slide

→ The following will be the cross-functional stakeholders teams:

- Sales/finance team
- Marketing team
- Data Analytics team
- Product development team (engineering)